YOLOSpecNN: A novel γ-ray spectra full-energy peak automatic search and segmentation model inspired by YOLO*

Cao-Lin Zhang,[1,2] Jiang-Mei Zhang,[1,2,†] Hao-Lin Liu,[1,2,3] Shu-Ya Qin,[4] and Jia-Qi Wang[1,2]

[1]*School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China*
[2]*CAEA Innovation Center of Nuclear Environmental Safety Technology,*
Southwest University of Science and Technology, Mianyang 621010, China
[3]*Department of Automation, University of Science and Technology of China, Hefei 230026, China*
[4]*Oxford Brookes College, Chengdu University of Technology, ChengDu 610059, China*

The qualitative identification and quantitative analysis of radioactive nuclides in unknown environments are essential for remote monitoring and prompt early warning of radioactive contamination. In recent years, deep learning techniques have made significant strides in automated qualitative identification. However, the quantitative analysis of radioactive nuclides still depends on traditional methods to determine peak positions and boundaries. These methods often require extensive manual expertise and parameter tuning, which makes it challenging to meet the demands of unmanned remote monitoring. This paper presents a novel framework for automatic full-energy peak segmentation, named YOLOSpecNN. We introduced a multi-Root Mean Square Error (RMSE) joint optimization function and developed a unified regression model capable of simultaneously predicting the central position, boundaries, and confidence of full-energy peaks. To address the challenge of low recall rates due to narrow, weak, and overlapping peaks, we proposed a new multi-scale context feature extraction module (MSNN module). This module effectively enhanced local detail features, significantly improving recall rates. The effectiveness of the proposed method was validated using six artificial radioactive nuclides (^{241}Am,^{57}Co,^{131}I,^{134}Cs,^{137}Cs, and ^{60}Co), along with ^{40}K, to construct a mixed energy spectrum dataset for quantitative evaluation. Experimental results show that the proposed method significantly outperforms traditional approaches, achieving a precision of 0.998, recall of 0.95, and the best F1 score of 0.974@0.427, and the average precision of 0.946. Compared to traditional morphological methods, the proposed method improves precision, recall, and the best F1 score by 0.512, 0.199, and 0.391, respectively. Ablation experiments further reveal that the MSNN module notably enhances recall, with an improvement of 0.067. Moreover, the proposed method performs excellently even in challenging environments with low gross counts and low Signal-to-Noise Ratio (SNR), achieving state-of-the-art (SOTA) results. Additionally, the model achieves an average real-time inference performance of 16.1941 ms on a 15 W low-power device. Overall, the proposed method demonstrates exceptional performance in the automatic search and segmentation of full-energy peaks, offering robust support for the implementation of unmanned remote radiation monitoring systems.

Keywords: Gamma Spectroscopy, Gamma-ray Spectral Analysis, Peak Searching and Segmentation, Interdisciplinary

I. INTRODUCTION

Gamma spectroscopy is crucial for analyzing gamma rays' energy peaks and intensities, identifying the characteristic energy of different radionuclides, and inferring their types and activities. This provides valuable data for nuclear physics research[1], environmental monitoring[2, 3], medical applications[4], and homeland security[5].

In gamma spectroscopy analysis, precisely defining the area of each full-energy peak in the γ-ray spectra is essential for accurately determining isotope energy positions and reducing background noise interference. Clear peak boundaries optimize the integration area, aiding in the calculation and quantitative analysis of peak areas. This enables researchers to calculate radioisotope radioactivity and assess the safety and environmental impact of samples.

However, challenges arise due to statistical fluctuations from low-count spectra's discrete Poisson fluctuations, low Signal-to-Noise Ratio (SNR), and peak overlap caused by limited energy resolution. These issues make automatic peak search and region segmentation difficult[6–8].

Gamma-ray spectrum analysis of radioisotopes generally involves two primary steps: peak search and peak area fitting[9, 10]. Early methods for peak search required smoothing to reduce statistical fluctuations[11], followed by the development of Fourier transform and wavelet-based techniques[12–16]. Recently, peak search methods based on image morphology have shown superior performance for complex full-energy peaks compared to traditional methods[17, 18]. However, factors such as the size of the morphological structural element, smoothing parameters, and false peak screening thresholds still require manual adjustment based on expert knowledge.

These methods heavily rely on human expertise to optimize parameter adjustments, balancing noise smoothing while retaining peak features. This reliance limits the automation of radioisotope identification devices, making them unsuitable for unmanned remote monitoring in communication-limited or harsh environments.

With the rapid development of Deep Learning (DL) technology, researchers have increasingly explored its potential for automatic feature extraction and radioisotope identification[19–21]. DL models have demonstrated high

accuracy in recognizing radionuclides[22, 23]. However, end-to-end multi-label classification models typically provide only qualitative probabilities for nuclide presence, without directly outputting the location and boundaries of characteristic peaks, which are often necessary for accurate quantitative analysis.

Recent studies have also explored neural networks for spectra analysis and peak detection[24, 25]. These methods directly output the location of characteristic peaks, offering a promising alternative to traditional radionuclide classification. However, accurate and automatic segmentation of full-energy peak regions remains unsolved, necessitating further fitting to determine optimal peak boundaries. Challenges such as peak stacking and shape distortion mean that parameters like fitting areas and peak shape functions still require expert intervention.

Fortunately, advancements in computer vision technology[26] offer similarities to our task, particularly in object detection and instance segmentation[27, 28]. These tasks aim to distinguish foreground from background, predict object regions, and identify individuals within the same category. Similarly, automatic separation of full-energy peaks from the background can be achieved. In computer vision, object detection typically involves predicting bounding boxes (x_1, x_2, y_1, y_2) and object category probabilities, effectively answering the question: "Where are the objects?"[27]. In this study, we aim to answer the question: "Where are the peaks?" The goal is to segment peak regions by learning two boundaries (x_1, x_2) along the energy axis from a 1D spectrum and distinguish peaks at various positions.

Inspired by the You Only Look Once (YOLO) object detection algorithm[29], we proposed a novel full-energy peak region segmentation model called YOLOSpecNN. This model predicts the center, width, and confidence of full-energy peaks in the energy spectrum, further calculating the left and right boundaries. To address challenges like weak, thin, and overlapping peaks, we introduced a multi-scale context feature extraction module that enhances local details and improves recall performance.

Using Monte Carlo simulations, we created a NaI(Tl) detector dataset for quantitative evaluation, incorporating factors such as gross counts, SNR, and nuclides mixing. The experimental results demonstrate the superior performance of our method.

II. MATERIALS AND METHODS

A. Backbone Network

This study adapts the ResNet-18 architecture [30] for one-dimensional implementation. The backbone network is composed of repeated residual blocks with varying output sizes [31]. Specifically, the proposed block structure incorporates skip connections, allowing the network to learn the residual mapping between input and output. This effectively mitigates issues such as vanishing and exploding gradients in deep neural network training.

This study retains only the convolutional layers from the original ResNet-18 model and replaces the fully connected layers with a Multi-Scale CNN (MSNN) module. The model input is a one-dimensional vector representing the energy spectrum, ($X \in \mathbb{R}^{1 \times 1024}$). The input energy spectrum is processed through the convolutional neural network to obtain the initial feature map (F_{resnet}), sized ($1 \times 64 \times 512$).

$$F_{resnet} = f_{resnet}(x; \theta_{resnet}), F_{resnet} \in \mathbb{R}^{(1 \times 64 \times 512)} \quad (1)$$

B. Multi-Scale CNN Module

In radionuclide spectra, some full-energy peaks cover large areas. Larger convolution kernels are intuitively required to capture larger spatial features. However, some radionuclide peaks are small in area (e.g., the full-energy peak of ^{241}Am occupies approximately 3/1024 of the spectrum), or mixed radionuclide spectra contain overlapping peaks with similar positions (e.g., 662 keV for ^{137}Cs and 605 keV for ^{134}Cs). In such cases, smaller convolution kernels are necessary to capture finer features.

The input spectrum is processed through the Backbone network to extract a deep and dense feature map, using a (1×3) convolution kernel, as in the original ResNet-18 architecture. This study introduces a Multi-Scale CNN Module (MSNN) for multi-scale convolutional feature fusion, designed to model the full-energy peak context of the spectrum, as shown in Fig. 1. By aggregating multi-scale context across different spectrum regions, the model can simultaneously capture both local fine features and global contextual information. This type of idea has been widely used in the field of computer vision. Contextual features refer to the spatial arrangement and relative positioning of pixels, which can enhance performance in areas like small object detection [32] and visual place recognition [33, 34].

The MSNN module uses four convolution kernels of different sizes: (1×1), (1×3), (1×5) and (1×7). These convolution kernels extract features from the input spectrum, generating feature maps of varying sizes, each with 128 channels. To align and aggregate feature maps at multiple scales, upsampling using nearest-neighbor interpolation is applied, resizing them to $(1 \times 64 \times 128)$. The feature maps are then concatenated, resulting in a multi-scale context-aggregated feature map (F_{MSNN}) with a size of $(1 \times 64 \times 512)$.

$$F_{MSNN} = f_{MSNN}(x; \theta_{MSNN}), F_{MSNN} \in \mathbb{R}^{(1 \times 64 \times 512)} \quad (2)$$

The final multi-scale feature map is flattened into a one-dimensional vector, followed by a fully connected layer. Finally, the linear output vector is resized to $(3 \times 1 \times S)$. If the center of the full-energy peak falls within a grid cell, that cell is responsible for detecting the object. In other words, the proposed model divides the input energy spectrum into $(1 \times S)$ grids.
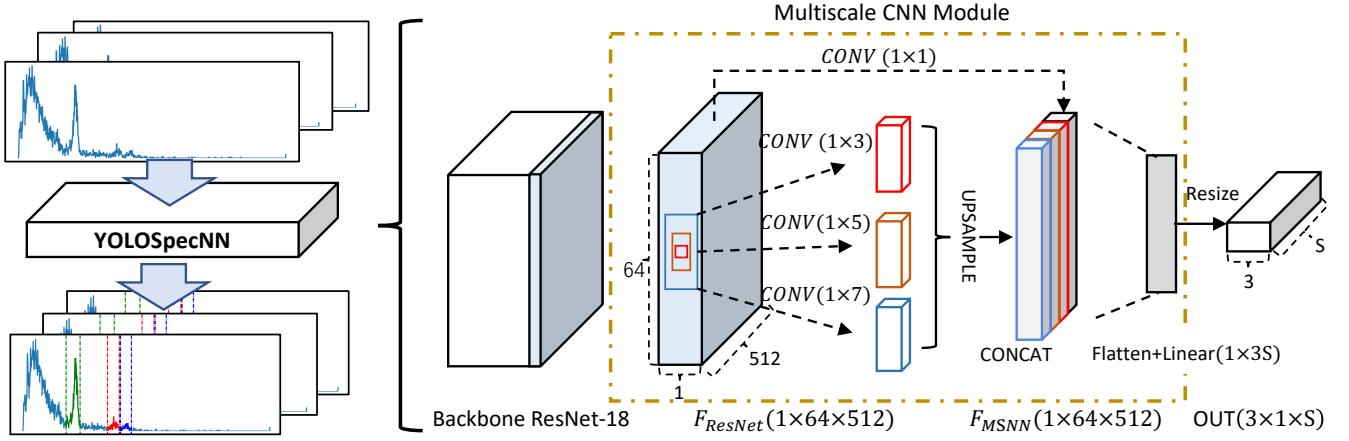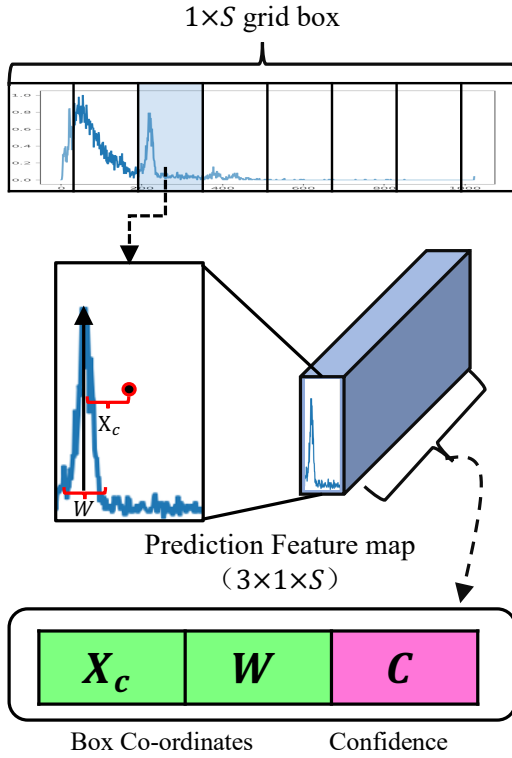
Fig. 1. Overall structural flow chart

$1 \times S$ grid box

Prediction Feature map
（$3 \times 1 \times S$）

X_c	W	C

Box Co-ordinates   Confidence

Fig. 2. Feature map predicted by the model

dicted relative to the entire spectrum. The confidence score (c) reflects the model's confidence in the presence of a peak within the box. If no peak is predicted in the cell, the confidence score should be zero. The final confidence prediction represents the Intersection over Union (IoU) between the predicted bounding box and the ground truth box.

The IoU measures the overlap between bounding boxes, calculated as the ratio of the intersection area to the union area of the two regions. In this study, the boundaries are one-dimensional coordinates, which differ from those in image-based domains. Therefore, the "area" in this context is expressed as the difference in coordinates. Specifically, A and B are the predicted boxes, with coordinate ranges $[x_1, x_3]$ and $[x_2, x_4]$, respectively. The IoU can be calculated as follows:

$$\begin{cases} A \cap B = x_3 - x_2 \\ A \cup B = x_4 - x_1 \end{cases} \quad (4)$$

$$IoU = \frac{A \cap B}{A \cup B} =$$

Fig. 3. IoU calculation schematic diagram

All output variables are constrained within the range $[0-1]$, achieved using the sigmoid function. Its mathematical representation is given by:

$$sigmoid\,(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The network's final output is a $(3 \times 1 \times S)$ tensor of predictions. In this study, $S = 384$, and the minimum resolution for the full-energy peak center interval is 3 channel bins.

$$S_{min} = ceil\,(1024/S) \quad (3)$$

As shown in Fig. 2, each grid cell predicts one bounding box and the associated confidence scores. Specifically, these are the center coordinate (x_c), width (w), and confidence score (c).

The coordinate (x_c) represents the center of the bounding box relative to the grid cell boundaries. The width (w) is pre-

4

C. Loss Function

During network training, three loss functions are jointly optimized for backpropagating the target boundary and confidence parameters, as shown in Equation 6.

$$\mathcal{L}_{total(x_i,w_i,C_i)} = \mathcal{L}_{objloc(x_i,w_i)} + \mathcal{L}_{objcon(C_i)} + \mathcal{L}_{noobjcon(C_i)} \tag{6}$$

The loss function expressions for the three components are provided in Equations (7)–(9):

$$\mathcal{L}_{objloc(x_i,w_i)} = \lambda_{coord} \sum_{i=0}^{S} \mathrm{i}_i^{obj} (x_i - \hat{x}_i)^2 \\ + \lambda_{coord} \sum_{i=0}^{S} \mathrm{i}_i^{obj} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 \tag{7}$$

$$\mathcal{L}_{objcon(C_i)} = \lambda_{coord} \sum_{i=0}^{S} \mathrm{i}_i^{obj} \left(C_i - \hat{C}_i\right)^2 \tag{8}$$

$$\mathcal{L}_{noobjcon(C_i)} = \lambda_{noobj} \sum_{i=0}^{S} \mathrm{i}_i^{noobj} \left(C_i - \hat{C}_i\right)^2 \tag{9}$$

$\mathcal{L}_{objloc(x_i,wi_i)}$ optimizes the predicted peak center coordinates (x_i) and width (w_i) to minimize the error between the predicted values and the labeled values $(\hat{x}_i)$ and $(\hat{w}_i)$. Additionally, to reduce the difference between large and small areas, we take the square root of the width(w).

$\mathcal{L}_{objcon(C_i)}$ minimizes the error between the predicted confidence (C_i) and the labeled confidence $(\hat{C}_i)$ when an object is present ($\mathrm{i}_i^{obj} = 1$) for the corresponding $bbox_i, i \in [0, S]$.

$\mathcal{L}_{noobjcon(C_i)}$ optimizes the confidence when no object is present ($i_i^{noobj} = 1$) in $bbox_i$.

We optimize the sum of squared errors in the model's output. Corresponding weights are used to adjust the contributions of positioning errors and confidence in non-target regions to the overall network error, thus preventing the gradient from being dominated by large noise baseline areas.

In this study, we increase the loss for bounding box coordinate prediction and reduce the loss for confidence prediction when no object is present. We use two parameters, λ_{coord} and λ_{noobj}, to achieve this, with $\lambda_{coord} = 1$ and $\lambda_{noobj} = 0.5$.

III. DATA PREPARATION

Various random factors, such as background conditions, measurement time, source intensity, and nuclide mixing, can affect the spectra of radionuclides measured in real-world environments. Large-scale measured datasets are challenging to acquire. We focus on three controllable variables: SNR, gross count, and nuclide mixing, to quantitatively evaluate our model. Monte Carlo simulations were conducted, and data augmentation was applied to address these variables.

A. NaI(Tl) Detector

NaI(Tl) detectors are widely used in gamma spectroscopy due to their high light yield, low background, low cost, ability to operate at room temperature, and ease of manufacturing in large sizes. However, their relatively low energy resolution can cause peak overlap, making automatic spectrum analysis challenging.

In this study, a homemade NaI(Tl) detector was modeled using Monte Carlo simulations, as shown in Figure 4. The detector consists of a 2-mm-thick stainless steel shell containing Fe, Cr, Ni, and C, a 3-inch cylindrical NaI(Tl) scintillator, a 0.3-mm-thick MgO reflective layer, and a SiPM photoelectric converter.
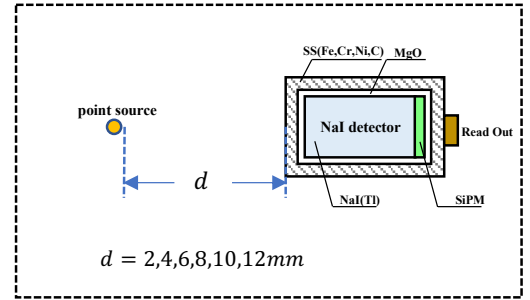

Fig. 4. Gamma detector structure and Monte Carlo simulation layout

B. Monte Carlo Simulation

Simulated spectra for six common artificial radionuclides (listed in Table 1) at various distances were generated using Geant4-11.1.3 [35]. The simulation model includes a point source with a gamma-ray energy range from 40 keV to 3000 keV. Each simulation involved the calculation of 100 million photons. We used FTFP_BERT as the reference physics list. The G4EmStandardPhysics was added to the physis list to simulate the physical process of photoelectric reaction. The energy deposition spectrum in the detector was obtained by inputting the nuclide's energy and branching ratio. The relative statistical uncertainty of the simulated results ranged from 0.0002 to 0.0016. Gaussian broadening was applied to simulate electronic noise.

$$FWHM(E_0) = C_1 + C_2\sqrt{E_0 + C_3 {E_0}^2} \tag{10}$$

$$E = \frac{FWHM}{2\sqrt{2ln2}}g + E_0 \tag{11}$$

Where E_0 is the energy deposition from the simulation, E is the broadened energy, $g \sim N(0,1)$ is a random variable following the standard normal distribution, and C_1, C_2 and C_3 are experimental coefficients, with values -0.0067, 0.0612, and -0.0451, respectively.

C. Data Augmentation

Data augmentation was performed using random gross count, SNR, and nuclide mixing ratios. These methods were widely used in previous studies and have proven effective in simulating statistical fluctuations and peak overlap in measured energy spectra [8, 22, 24].

Random gross count and Random SNR: Random gross count (C_{gross}) is achieved by sampling the spectrum as a probability distribution across energy.

$$C_{gross} = C_{source} + C_{bq} \tag{12}$$

$$SNR = C_{source}/\left(C_{source} + C_{bq}\right) \tag{13}$$

Given a randomly selected low SNR, we calculated the background count (C_{bq}) and nuclide count (C_{source}) through sampling. A background spectrum was obtained by measuring the shielded detector with lead bricks for 3600 seconds, which included natural radiation from the ^{40}K isotope. The simulated spectrum is then linearly combined with the measured background spectrum for different SNR conditions.

Nuclide mixing and Superposition: The count of each nuclide was obtained separately based on gross count, SNR, and random proportion, then linearly superimposed. The category of mixed nuclides was generated through permutation and combination. Assuming there are n mixed nuclides, n ratios k_n are randomly generated. To prevent large variations in nuclide counts, the initial mixed ratio range was set between 0.1 and 0.5.

$$K = \{k_1, k_2, k_3 \ldots k_n\}, k_n \sim unif\left([0.1, 0.5]\right) \tag{14}$$

Then, normalize K values to sum 1:

$$K_{norm} = K/sum\left(K\right) \tag{15}$$

According to the gross count C_{source} required by the random SNR, the count C_s of each nuclide source is obtained:

$$C_s = \{c_{s1}, c_{s2}, c_{s3} \ldots c_{sn}\} = C_{source} \times K_{norm} \tag{16}$$

We can perform data augmentation steps based on the required number of samples to meet the needs of training and performance evaluation.

Fig. 5 shows the comparison between the simulated synthetic spectrum (orange) and the measured spectrum (blue). For the simulated spectra, we manually adjusted the counts of ^{137}Cs and background to 3e3, and the counts of ^{60}Co to 1e3. We used the detector with the same structure as Fig. 4 and obtained the measured energy spectrum based on ^{137}Cs and ^{60}Co point sources. We calculated the cosine similarity between the simulated synthetic spectrum and the measured
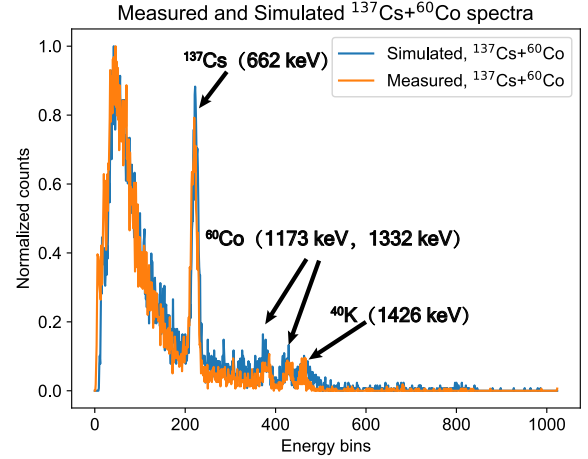

Fig. 5. Measured and Simulated ^{137}Cs+^{60}Co mixed spectra

spectrum, which is 0.965, indicating a good fit between the two.

The constructed dataset generally includes six common artificial radionuclides and the natural radionuclide ^{40}K. Among them, ^{241}Am and ^{57}Co have low energy, and their full-energy peaks occupy a small area, which overlaps with the strong low-energy Compton scattering region. ^{131}I, ^{60}Co, ^{134}Cs, and ^{137}Cs are common materials released in nuclear power plant accidents [36–38]. When they coexist with the natural radionuclide background ^{40}K, full-energy peak overlap may occur.

For each nuclide category, the center position (x_c), area width (w), and corresponding confidence (c) of the full-energy peaks will be labeled in all spectral samples to construct the dataset used in this study. All energy spectra are normalized to ensure that the input depends solely on the shape.

IV. EVALUATION

A. Evaluation Metrics

Evaluation methods widely used in computer vision object detection models, including precision (P), recall (R) and F1 score (F1), and average precision (AP), were used to evaluate our method [39, 40].

Recall: The percentage of correctly detected targets relative to the total number of targets.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

Precision: The proportion of correctly detected targets relative to the total predicted targets.

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

Table 1. Data enhancement parameter table.

Radioisotope	SNR	Gross Counts	Mixed quantity	Initial mixed ratio
^{241}Am,^{57}Co,^{131}I, ^{134}Cs,^{137}Cs,^{60}Co,^{40}K	$0.1 \sim 1$	$1e3 \sim 1e5$	$1 \sim 5\ Artificial\ Nuclides +^{40}K$	$0.1 \sim 0.5$

F1 score: The harmonic mean of precision and recall, which evaluates model performance in a balanced manner, considering both false positives and false negatives.

$$F1 = \frac{2TP}{2TP + FP + FN} = 2\frac{Precision \times Recall}{Precision + Recall} \quad (19)$$

Here, TP denotes the number of correctly identified targets, FP indicates the number of falsely detected targets, and FN refers to the number of undetected targets.

Average Precision (AP): A measure of the area under the precision-recall (P-R) curve, with recall on the x-axis and precision on the y-axis. AP effectively evaluates the average detection precision across varying recall levels. The calculation formula for AP is shown in equation 20.

$$AP = \int_0^1 P_i\left(R_i\right) d_{R_i} \quad (20)$$

B. Compared Methods

Morphological-based: Rui Luo et al.[17] compared five traditional peak search methods with a new morphology-based method using mixed spectra generated from multiple nuclides. The results indicated that the morphological method is more suitable for peak search in mixed spectra generated from multiple nuclides. Therefore, this method serves as the baseline for comparison in this paper.

The morphological peak search method consists of three steps: spectrum smoothing, morphological top-hat transformation, and false peak screening. Finally, significant peaks in the energy spectrum are retained. Spectrum smoothing was performed using the Savitzky-Golay filter[41]. The parameters to adjust are the window size (w) and the polynomial order (k). As w increases and k decreases, the smoothing effect becomes more pronounced. In the morphological top-hat transformation, the structural element size (L) must be adjusted. If L is too small, excessive noise will be detected. If L is too large, the algorithm's ability to distinguish adjacent peaks will decrease. After the morphological white-hat transformation, the spectrum was divided into continuous non-zero regions.

The width (w_i) of each region was then calculated, and the threshold (W_T) was set. If $w_i < W_T$, it is classified as a false peak. Multiple candidate peaks may be detected. When the difference between the detected peak and the ground truth is within the energy window (set to 2% in this paper), the search is deemed correct, and others are classified as false peaks [42].

In total, four parameters must be manually adjusted. The adjusted parameters and their ranges are summarized in Table

Table 2. Parameter table for morphological-based method.

Steps	Smoothing		Top hat	False peak screening
Parameters	(w)	(k)	(L)	W_T
Traversal range	9~49	3,5,7	4~40	2~15
Best parameters	25	5	16	14

2. By iterating over different parameters and using F1 as the evaluation metric, the best parameter configuration is identified, as shown in Table 2. In the subsequent testing, it will be labeled to as **MorPh**.

ResNet-18: To evaluate the effectiveness of the proposed MSNN module, an ablation test was conducted. In the implementation, to ensure experimental fairness, the MSNN module was removed from the architecture shown in Fig. 1. The feature map F_{resnet} was then directly expanded and connected to the following fully connected layer, labeled as **ResNet-18** in the test. The method proposed in this paper is labeled as **ResNet-18+MSNN**.

C. Model Training

During the model training process, the dataset was split into an 80% training set and a 20% validation set. The initial learning rate was set to Ir $= 1 \times 10^{-4}$, and the batch size was set to 256. The learning rate decayed by a factor of 10 every 30 epochs. The Adam optimizer was used for parameter optimization.

It is worth noting that during each data read in the training process, uniformly distributed noise in the range of $0.001 \sim 0.02$ and peak displacements in the range of $0 \sim 200$ were randomly added to further augment the dataset and ensure its diversity. These two data augmentation steps were only applied during training, ensuring that the test dataset remained exclusive and never overlapped with the training data.

Fig. 6(a) shows the trend of the total loss curves during the training process. Fig. 6(b) shows the performance of various indicators on the validation data set, showing an upward trend corresponding to the Loss.

D. Evaluation Results and Analysis

A comprehensive performance evaluation has been first performed in this paper, followed by the control of gross count, SNR, and nuclide category for quantitative analysis. Subsequently, qualitative evaluation and inference time performance have been assessed.

Comprehensive performance: Values for the SNR and gross count ware randomly selected based on the range specified in Table 1, and a test set of 60,000 samples is generated.

Table 3. Comprehensive performance comparison table.

Method	Performance metrics			
	best F1@confidence	Precision	Recall	AP
MorPh	0.583	0.477	0.751	-
ResNet-18	0.935@0.629	0.994	0.883	0.86
ResNet-18+MSNN	**0.974@0.427**	**0.998**	**0.95**	**0.946**
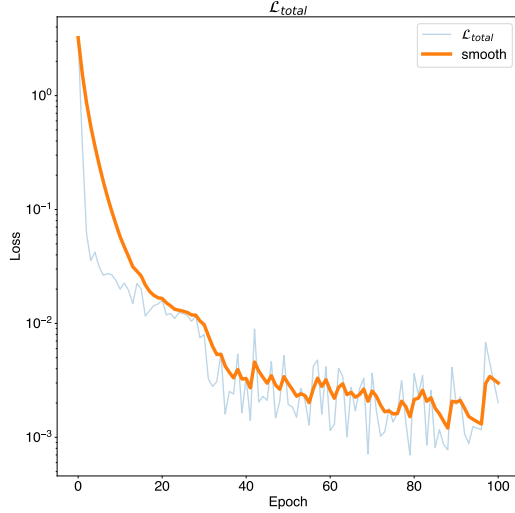
(a) Loss curve during training
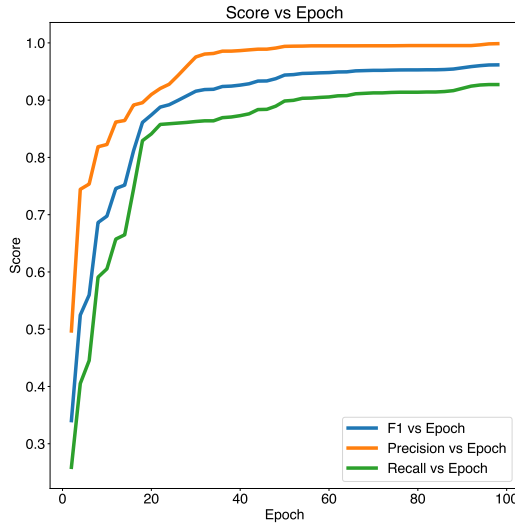
(b) Performance of validation dataset

Fig. 6. Performance of validation dataset during training

Figures 7 and Table 3 present the recall and precision performance of the three methods on the test set.

The DL-based method calculates the corresponding recall and precision at different confidence thresholds and generates a P-R curve. The morphology-based method visualizes the calculation results during the parameter iteration process as a scatter plot. The DL-based method outperforms the morphology-based method in both recall and precision, particularly in terms of high precision.

Compared to the morphology-based method, ResNet-18's precision, recall rate, and best F1 score improved by 0.517, 0.082, and 0.352, respectively. After implementing the MSNN module, these values increased by 0.521, 0.199, and 0.391, respectively. Table 3 shows that, compared to the ResNet-18 model, the AP improves by 0.072 after implementing the MSNN module. Furthermore, Figure 7(b) shows that after implementing the MSNN module, the optimal confidence threshold is lower, which typically results in a higher recall rate. Therefore, the results indicate that the main advantage of the MSNN module is the improvement in recall rate.

Low gross count: The peaks in the low-count spectrum are influenced by statistical fluctuations, which are affected by factors such as the radiation source intensity, detector efficiency, and measurement time. This results in the misidentification of numerous false peaks. A stronger detection capability of peaks in the low-count spectrum indicates that the proposed algorithm has enhanced anti-noise ability and can analyze low-activity radionuclide spectra more efficiently.

The fixed gross counts were set to 1e3, 2e3, 4e3, 6e3, 8e3, and 1e4, respectively. To reduce the impact of low SNR, the random SNR was controlled within the range of $0.5 \sim 1$, and 20,000 samples were regenerated for performance testing. The best F1 score, precision, and recall of the three methods were evaluated based on the test set.

As shown in Fig. 8, as the gross count decreases, the performance indicators of all three methods also decrease. However, the F1 score of the ResNet-18+MSNN model outperforms the other two methods, as shown in Fig. 8(a).

As shown in Fig. 8(c), the recall of the morphology-based method is less affected and even matches that of the DL-based method. That's because large statistical fluctuations can increase the false alarm rate, but the method can still identify the existing peaks.

The increased false alarm rate leads to a rapid decrease in precision, as shown in Fig. 8(b). As shown in Fig. 8(c), the recall rate of the method without the MSNN module gradually decreases as statistical fluctuations increase.

Overall, as shown in Fig. 8(b), the DL-based method consistently maintains its advantage in high precision. The MSNN module effectively improves the recall rate.

Low SNR: The SNR of the spectrum indicates the prominence of the peak relative to the background, which is typically influenced by the intensity of the radioactive source's background. A stronger peak detection capability in low SNR spectra suggests that the proposed algorithm can effectively identify weak peaks buried in a strong background.
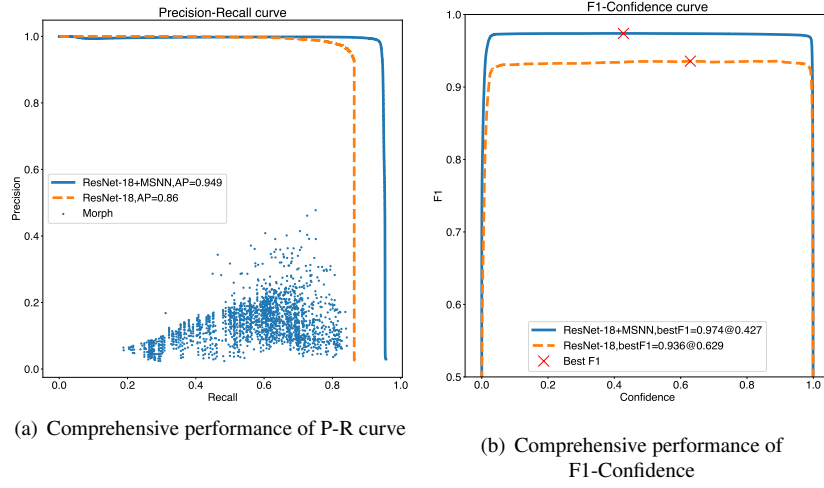
To evaluate performance under low SNR conditions, the
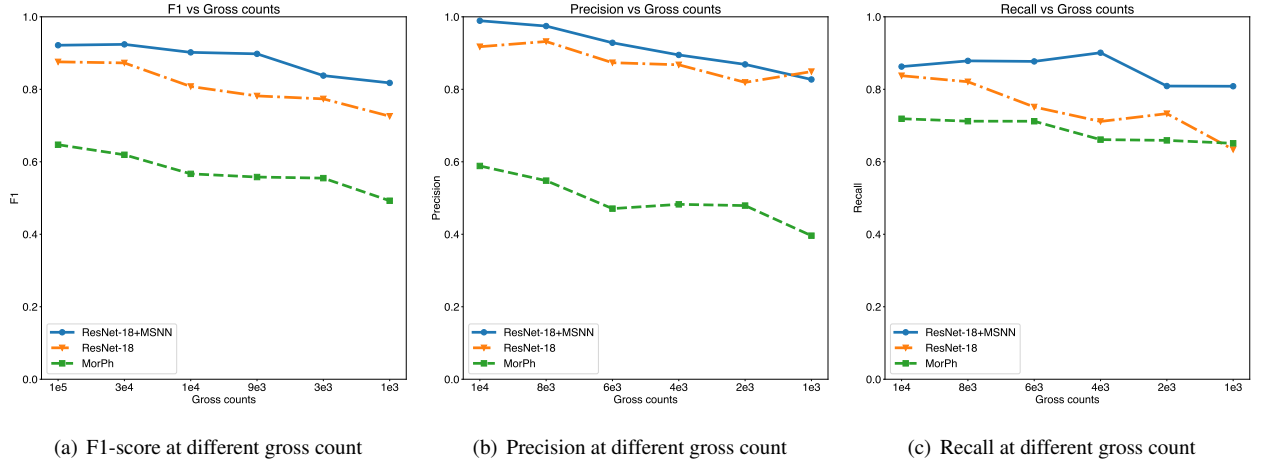
(a) Comprehensive performance of P-R curve

(b) Comprehensive performance of F1-Confidence

Fig. 7. Comprehensive performance comparison chart

(a) F1-score at different gross count

(b) Precision at different gross count

(c) Recall at different gross count

Fig. 8. Performance of different low gross count

(a) F1-score at different SNR

(b) Precision at different SNR

(c) Recall at different SNR
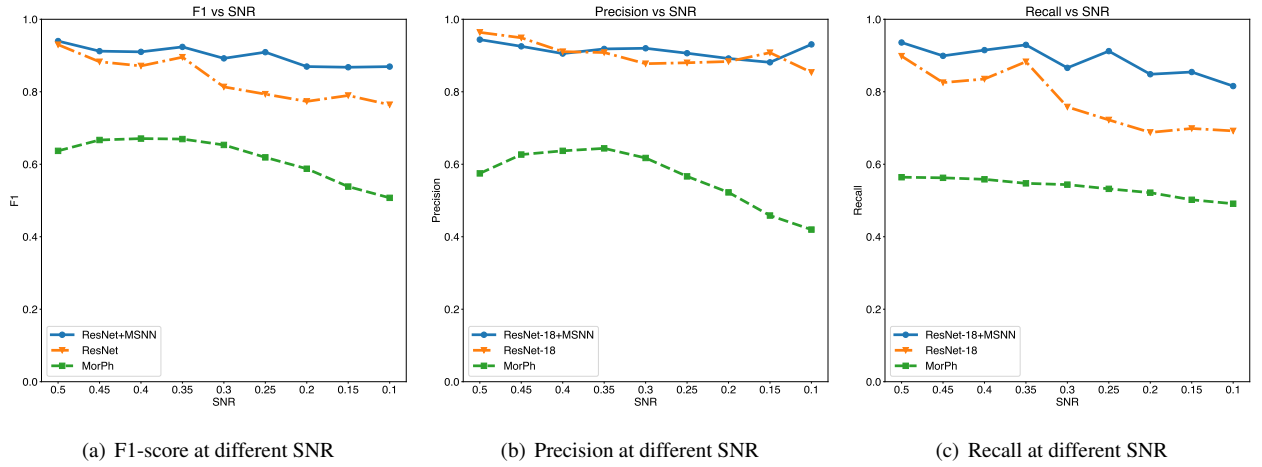
Fig. 9. Performance of different low SNR

460 SNR values ware set to 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 461 0.45, and 0.5. To minimize the impact of low gross counts, the

(a) F1-score at different radionuclide

(b) Precision at different radionuclide
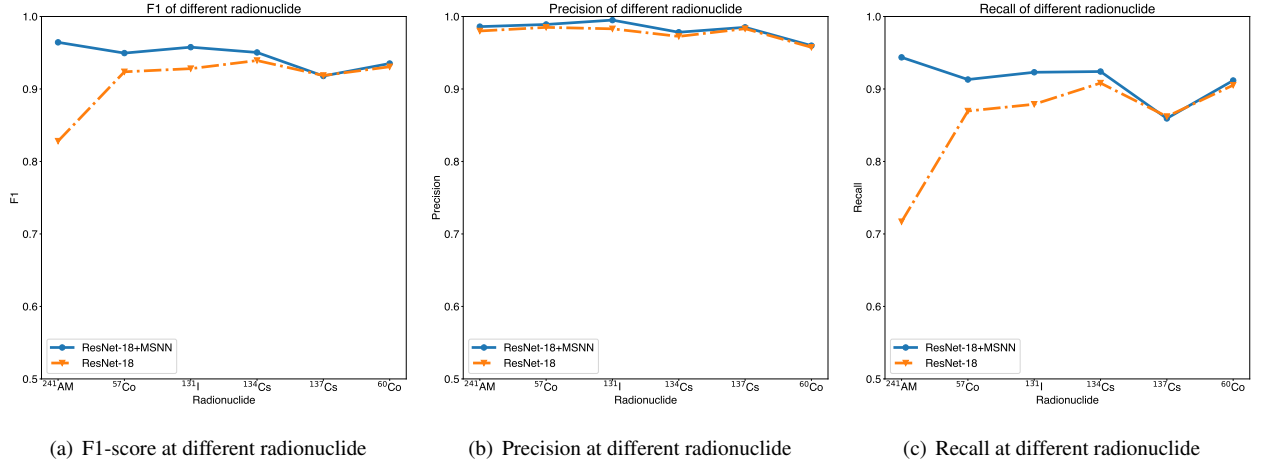
(c) Recall at different radionuclide

Fig. 10. Performance of different radionuclide

random gross count range was controlled between $4e4 \sim 1e5$, and 20,000 samples are regenerated for performance testing. The impact of SNR on the three methods was evaluated, as shown in Fig. 9.

The DL-based method consistently maintains higher precision across all SNR conditions. Although the F1 score decreases at low SNR, it remains higher than that of the morphology-based method. As shown in Fig. 9(c), the recall rate is significantly improved by the MSNN module, particularly when the SNR is low ($SNR < 0.3$).

Radionuclide category: Different radionuclide categories can represent peaks with different half-widths to a certain extent. A smaller FWHM corresponds to a smaller peak area. Small object detection remains a challenging task in computer vision [43]. Previous experimental results showed that the MSNN module improves the recall rate more significantly than the precision. To further investigate the reasons, this paper separately evaluates the detection performance of DL-based methods for peaks of varying widths.

The test set consists of samples containing only one type of radionuclide. To reduce the impact of low gross count and low SNR, the random gross count range was set to $5e4 \sim 1e5$, the SNR range to $0.5 \sim 1.0$, and 20,000 samples were regenerated for performance testing.

Fig. 10 shows the performance of the DL-based method for different radionuclides, evaluating the model's performance for thinner and wider peaks. As shown in Fig. 10(c), without the MSNN module, the recall rate for the ^{241}Am and ^{57}Co peaks is significantly lower than that for other nuclides. This suggests that the missed detection rate is higher due to the small area occupied by thin peaks. After applying the MSNN module, the recall rate improves significantly, demonstrating the effectiveness of the proposed method.

Additionally, as shown in Fig. 10(c), the recall rate for ^{137}Cs is significantly lower than for ^{134}Cs and ^{60}Co, despite the similar FWHM of their peaks. This may be attributed to the small number of characteristic peaks of ^{137}Cs. In other words, the effective area occupied by the characteristic peak of ^{137}Cs in the complete energy spectrum is small, and the

number of positive and negative samples is highly imbalanced.

Qualitative analysis: Fig. 11 to Fig. 12 illustrate the peak search and segmentation results of the morphological method, ResNet-18, and ResNet-18+MSNN. The morphological peak search results are marked with red arrows. The DL model output includes confidence (blue bar graph), characteristic peak area (red rectangle), and the ground truth (gt) value, indicated by a green dotted rectangle.

As shown in Fig. 11, the energy spectrum contains mixed nuclides of ^{57}Co, ^{60}Co, ^{134}Cs, and ^{137}Cs, with significant overall statistical fluctuations, and overlapping characteristic peaks at 605 keV (^{134}Cs) and 662 keV (^{137}Cs).

In this case, the morphological method generates numerous false positives, leading to reduced precision. However, the ResNet-18 model does not provide sufficient confidence for the 605 keV (^{134}Cs) characteristic peak, resulting in a lower recall rate. The segmentation precision of the ResNet-18 model for the 1173 keV (^{60}Co) characteristic peak area is significantly lower than that of the ResNet-18+MSNN model.

The energy spectrum shown in Fig. 12 contains mixed nuclides of ^{241}Am, ^{57}Co, ^{60}Co, ^{134}Cs, and ^{137}Cs. The statistical fluctuation of this energy spectrum sample is relatively small, with challenges posed by thin peaks (59keV of ^{241}Am) and overlapping peaks (605keV of ^{134}Cs and 662keV of ^{137}Cs).

The morphology-based method has a lower false alarm rate. However, additional false peak screening or manual parameter adjustments are still required.

The DL-based method retains its high precision advantage. However, the ResNet-18 model overlooks the characteristic peak of ^{241}Am, and the confidence prediction of the 605kev characteristic peak of ^{134}Cs is worse than that of ResNet-18+MSNN.

The morphological method inevitably generates numerous false peaks due to the interference of strong statistical fluctuations in low-count full-energy peaks. The DL-based method, with its strong nonlinear fitting ability, effectively handles noise interference, maintaining its high precision advantage. After applying the MSNN module, the intricate features of
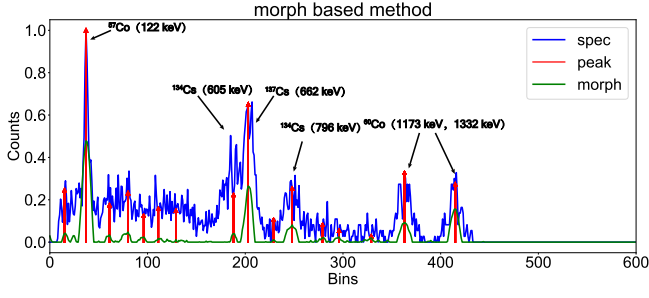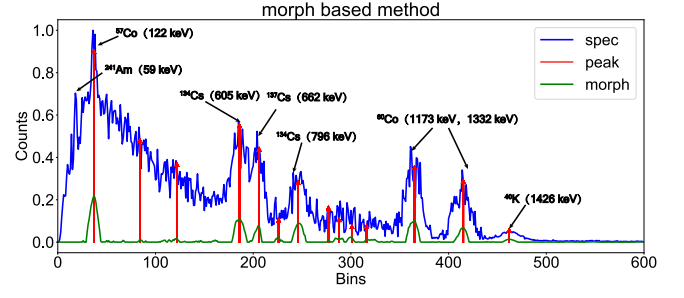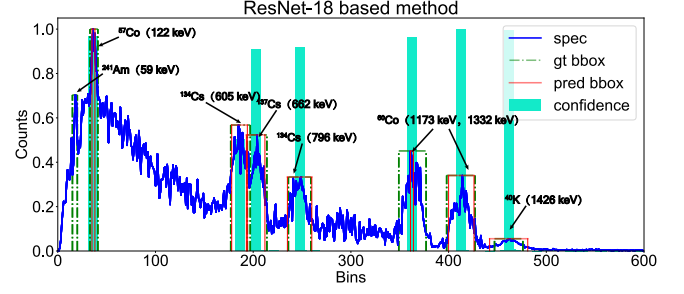
(a) Morph based method
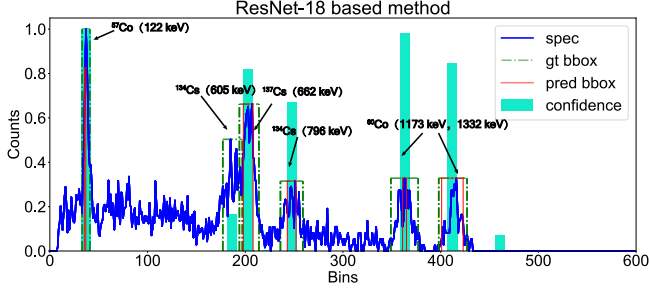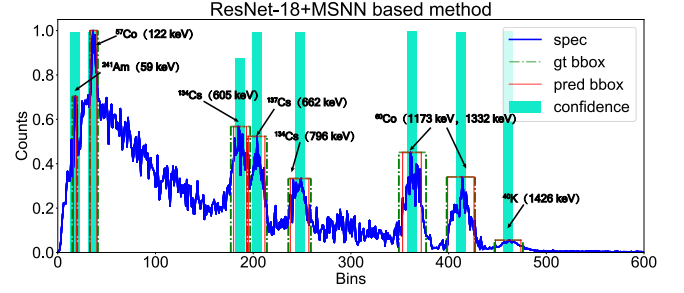

(b) ResNet-18 based method
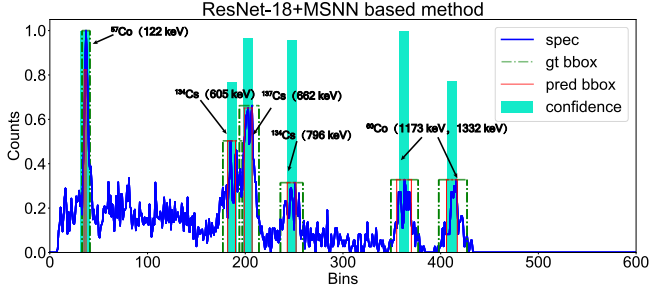
(c) ResNet-18 + MSNN based method

Fig. 11. Qualitative analysis of ^{57}Co, ^{60}Co, ^{134}Cs, and ^{137}Cs mixed Radionuclides

(a) Morph based method

(b) ResNet-18 based method

(c) ResNet-18 + MSNN based method

Fig. 12. Qualitative analysis of ^{241}AM, ^{57}Co, ^{60}Co, ^{134}Cs, and ^{137}Cs mixed Radionuclides

the overlapping peak area can be effectively modeled, further improving the recall rate.

Inference time: Real-time performance is a critical indicator in practical monitoring systems. The NVIDIA Jetson AGX Orin 32GB module was used to evaluate inference time performance, allowing users to set power consumption limits of 15W, 30W, 50W, and 60W. This setup enables the evaluation of both inference time and power consumption under various application scenarios.

We conducted experiments on 60,000 samples to measure inference times under various power consumption constraints, using Python 3.8 and PyTorch 2.1. The results are presented in Table 4.

As shown in Table 4, under a 15 W power consumption limit, the average inference time per sample is 16.1941 ms, with a maximum of 112.4730 ms and a minimum of 13.5004 ms, demonstrating the good real-time performance of the proposed model.

Table 4. Inference time performance.

Power(W)	Inference time (ms)		
	min	average	max
15	13.5004	16.1941	112.4730
30	9.5539	10.4092	30.1266
50	10.3586	11.4204	21.4953
60	7.2591	7.8072	21.9126

V. CONCLUSIONS

This paper has presented YOLOspecNN, a novel method that applies computer vision object detection techniques to radionuclide spectrum peak search and segmentation. Our method automatically segments the full-energy peak regions of gamma spectra without manual parameter adjustments, while supporting both qualitative identification and quantitative analysis.

The proposed approach offers significant advantages over traditional morphology-based methods, particularly in handling challenges such as complex mixed nuclide energy spectra, low gross counts, low SNR, and overlapping peaks. Additionally, we have introduced an MSNN module that enhances local context features, improving recall performance for weak peaks or thin peaks in low-energy regions. The YOLOspecNN model performs real-time inference on a 15W low-power embedded platform.

Full-energy peaks exhibit relatively fixed Poisson or Gaussian statistical properties, which increases the model's flexibility and may enable better handling of various detectors and spectra. In the future, we plan to expand the dataset to include additional detector types and nuclide categories, further enhancing the model's robustness and generalization.

[1] J. Eberth, J. Simpson, From Ge(Li) detectors to gamma-ray tracking arrays-50 years of gamma spectroscopy with germanium detectors. Prog. Part. Nucl. Phys. **60**, 283-337 (2008). https://doi.org/10.1016/j.ppnp.2007.09.001

[2] J. Klusoň, Environmental monitoring and in situ gamma spectrometry. Radiat. Phys. Chem. **61**, 209-216 (2001). https://doi.org/10.1016/S0969-806X(01)00242-0

[3] H. Kofuji, In situ measurement of ^{134}Cs and ^{137}Cs in seabed using underwater γ-spectrometry systems: application in surveys following the Fukushima Dai-ichi Nuclear Power Plant accident. J. Radioanal. Nucl. Chem. **303**, 1575-1579 (2015). https://doi.org/10.1007/s10967-014-3702-0

[4] Y. Eisen, A. Shor, I. Mardor, CdTe and CdZnTe X-ray and gamma-ray detectors for imaging systems. IEEE. Trans. Nucl. Sci. **51**, 1191-1198 (2004). https://doi.org/10.1109/TNS.2004.829437

[5] D. Fagan, S. Robinson, R. Runkle, Statistical methods applied to gamma-ray spectroscopy algorithms in nuclear security missions. Appl. Radiat. Isot. **70**, 2428-2439 (2012). https://doi.org/10.1016/j.apradiso.2012.06.016

[6] M. Alamaniotis, S. Lee, T. Jevremovic, Intelligent Analysis of Low-Count Scintillation Spectra Using Support Vector Regression and Fuzzy Logic. Nucl. Tech. **191**, 41-57 (2015). https://doi.org/10.13182/NT14-75

[7] M. Monterial, K. Nelson, S. Labov et al., Benchmarking Algorithm for Radio Nuclide Identification (BARNI) Literature Review. (2019,2). https://www.osti.gov/biblio/1544518

[8] S. Qi, S. Wang, Y. Chen et al., Radionuclide identification method for NaI low-count gamma-ray spectra using artificial neural network. Nucl. Eng. Technol. **54**, 269-274 (2022). https://doi.org/10.1016/j.net.2021.07.025

[9] M. Rawool-Sullivan, J. Bounds, S. Brumby, et al., Steps toward automated gamma ray spectroscopy. Los Alamos National Laboratory (2010)

[10] K. Lam, W. Zhang, Gamma peak search and peak fitting algorithm for a low-resolution detector with applications in gamma spectroscopy. J. Radioanal. Nucl. Chem. **322**, 255-261 (2019,11). https://doi.org/10.1007/s10967-019-06760-x

[11] J. Routti, S. Prussin, Photopeak method for the computer analysis of γ-ray spectra from semiconductor detectors. Nucl. Instrum. Methods. **72**, 125-142 (1969). https://doi.org/10.1016/0029-554X(69)90148-7

[12] I.A. Slavić, S.P. Bingulac, A simple method for full automatic gamma-ray spectra analysis. Nucl. Instrum. Methods. **84**, 261-268 (1970). https://doi.org/10.1016/0029-554X(70)90270-3

[13] G. Xiao, L. Deng, B. Zhang et al., A nonlinear wavelet method for data smoothing of low-level gamma-ray spectra. J. Nucl. Sci. Technol **41**, 73-76 (2004). https://doi.org/10.3327/jnst.41.73

[14] C.J. Sullivan, S.E. Garner, K.B. Butterfield, Wavelet analysis of gamma-ray spectra. *IEEE Symposium Conference Record Nuclear Science 2004*. **1** pp. 281-286 Vol. 1 (2004). https://doi.org/10.1109/NSSMIC.2004.1462198

[15] C.J. Sullivan, M.E. Martinez, S.E. Garner, Wavelet analysis of sodium iodide spectra. *IEEE Nuclear Science Symposium Conference Record*, 2005 **1** pp. 302-306 (2005). https://doi.org/10.1109/NSSMIC.2005.1596258

[16] H. Yang, X. Zhang, W. Gu et al., A novel method for gamma spectrum analysis of low-level and intermediate-level radioactive waste. Nucl. Sci. Tech. **34**, 87 (2023). https://doi.org/10.1007/s41365-023-01236-w

[17] R. Luo, Y. Huang, X. Liu, Comparative study on peak-seeking methods of the mixed radioactive energy spectrum. Open Access Library Journal **7**, 1-10 (2020). https://doi.org/10.4236/oalib.1106676

[18] Y. Huang, M. Liu, X. Liu et al., Nuclide spectrum peak searching algorithm based on multiple morphological structuring elements. Journal Of Physics: Conference Series **1634**, 012078 (2020). https://doi.org/10.1088/1742-6596/1634/1/012078

[19] H. Liu, H. Ji, J. Zhang et al., Novel algorithm for detection and identification of radioactive materials in an urban environment. Nucl. Sci. Tech. **34**, 154 (2023). https://doi.org/10.1007/s41365-023-01304-1

[20] H. Liu, H. Ji, J. Zhang et al., A novel approach for feature extraction from a gamma-ray energy spectrum based on image descriptor transferring for radionuclide identification. Nucl. Sci. Tech. **33**, 158 (2022). https://doi.org/10.1007/s41365-022-01150-7

[21] J. Wang, W. Gu, H. Yang et al., Analytical method for γ energy spectrum of radioactive waste drum based on deep neural network. Nucl. Tech. 45, 040501 (2022). https://doi.org/10.11889/j.0253-3219.2022.hjs.45.040501 **(in Chinese)**

[22] S. Qi, W. Zhao, Y. Chen et al., Comparison of machine learning approaches for radioisotope identification using NaI(TI) gamma-ray spectrum. Appl. Radiat. Isot. **186** pp. 110212 (2022). https://doi.org/10.1016/j.apradiso.2022.110212

[23] S. Galib, P. Bhowmik, A. Avachat et al., A comparative study of machine learning methods for automated identification of radioisotopes using NaI gamma-ray spectra. Nucl. Eng. Technol. **53**, 4072-4079 (2021). https://doi.org/10.1016/j.net.2021.06.020

[24] S. Wu, X. Tang, P. Gong et al., Peak-searching method for low count rate γ spectrum under short-time measurement based on a generative adversarial network. Nucl. Instrum. Methods Phys. Res. A. **1002** pp. 165262 (2021). https://doi.org/10.1016/j.nima.2021.165262

[25] R. Zhao, N. Liu, Low-resolution gamma-ray spectrum analysis using comprehensive training set and deep ResNet architecture. Nucl. Instrum. Methods Phys. Res. A. **1050** pp. 168135 (2023). https://doi.org/10.1016/j.nima.2023.168135

[26] A. Voulodimos, N. Doulamis, A. Doulamis et al., Deep Learning for Computer Vision: A Brief Review. Intell. Neuroscience

2018 (2018,1). https://doi.org/10.1155/2018/7068349

[27] Z. Zou, K. Chen, Z. Shi, et al., Object Detection in 20 Years: A Survey. Proceedings Of The IEEE **111**, 257-276 (2023). https://doi.org/10.1109/JPROC.2023.3238524

[28] A. Hafiz, G. Bhat, A survey on instance segmentation: state of the art. Int. J. Multimed. Inf. Retr. **9**, 171-189 (2020,9). https://doi.org/10.1007/s13735-020-00195-x

[29] J. Redmon, S. Divvala, R. Girshick et al., You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)* pp. 779-788 (2016). https://doi.org/10.1109/CVPR.2016.91

[30] K. He, X. Zhang, S. Ren et al., Deep Residual Learning for Image Recognition. *2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)* pp. 770-778 (2016). https://doi.org/10.1109/CVPR.2016.90

[31] K. He, X. Zhang, S. Ren et al., Identity Mappings in Deep Residual Networks. Computer Vision – ECCV 2016 pp. 630-645 (2016)

[32] G. Qi, Y. Zhang, K. Wang et al., Small Object Detection Method Based on Adaptive Spatial Parallel Convolution and Fast Multi-Scale Fusion. Remote. Sens-Basel. **14** (2022). https://doi.org/10.3390/rs14020420

[33] Z. Xin, Y. Cai, T. Lu et al., Localizing Discriminative Visual Landmarks for Place Recognition. *2019 International Conference On Robotics And Automation (ICRA)* pp. 5979-5985 (2019). https://doi.org/10.1109/ICRA.2019.8794383

[34] H. Kim, E. Dunn, J. Frahm, Learned Contextual Feature Reweighting for Image Geo-Localization. *2017 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)* pp. 3251-3260 (2017). https://doi.org/10.1109/CVPR.2017.346

[35] S. Agostinelli, J. Allison, K. Amako et al., Geant4—a simulation toolkit. Nucl. Instrum. Methods Phys. Res. A. **506**, 250-303 (2003), https://doi.org/10.1016/S0168-9002(03)01368-8

[36] J. Li, S. Liu, Y. Zhang et al., Pre-assessment of dose rates of ^{134}Cs, ^{137}Cs, and ^{60}Co for marine biota from discharge of Haiyang Nuclear Power Plant, China. J. Environ. Radioactiv. **147** pp. 8-13 (2015). https://doi.org/10.1016/j.jenvrad.2015.05.001

[37] S. Ueda, H. Hasegawa, H. Kakiuchi et al., Fluvial discharges of radiocaesium from watersheds contaminated by the Fukushima Dai-ichi Nuclear Power Plant accident, Japan. J. Environ. Radioactiv. **118** pp. 96-104 (2013). https://doi.org/10.1016/j.jenvrad.2012.11.009

[38] G. Katata, M. Ota, H. Terada et al., Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source term estimation and local-scale atmospheric dispersion in early phase of the accident. J. Environ. Radioactiv. **109** pp. 103-113 (2012). https://doi.org/10.1016/j.jenvrad.2012.02.006

[39] R. Padilla, S. Netto, E. Silva, A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference On Systems, Signals And Image Processing (IWSSIP)* pp. 237-242 (2020). https://doi.org/10.1109/IWSSIP48289.2020.9145130

[40] R. Padilla, W. Passos, T. Dias et al., A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. Electronics-Switz. **10** (2021). https://doi.org/10.3390/electronics10030279

[41] H. Yule, Mathematical smoothing of gamma ray spectra. Nucl. Instrum. Methods. **54**, 61-65 (1967). https://doi.org/10.1016/S0029-554X(67)80007-7

[42] S. Islami rad, Optimization of energy window for gamma densitometer based backscatter method in oil industry. Russ. J. Nondestruct+. **52**, 245-249 (2016,4). https://doi.org/10.1134/S1061830916040045

[43] Y. Liu, P. Sun, Wergeles, N et al., A survey and performance evaluation of deep learning methods for small object detection. Expert. syst. Appl. **172** pp. 114602 (2021). https://doi.org/10.1016/j.eswa.2021.114602